A Hybrid Recommender System to Improve Circular Economy in Industrial Symbiotic Networks

Anna Gatzioura [1,2], **Miquel Sànchez-Marrè [1,2,*]** and **Karina Gibert [1,3]**

[1] Knowledge Engineering & Machine Learning Group at Intelligent Data Science and Artificial Intelligence Research Centre (KEMLG-@-IDEAI), Universitat Politècnica de Catalunya BarcelonaTech (UPC), Catalonia, 08034 Barcelona, Spain; anna.gatzioura@gmail.com (A.G.); karina.gibert@upc.edu (K.G.)

[2] Department of Computer Science, Universitat Politècnica de Catalunya BarcelonaTech (UPC), Catalonia, 08034 Barcelona, Spain

[3] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya BarcelonaTech (UPC), Catalonia, 08034 Barcelona, Spain

* Correspondence: miquel@cs.upc.edu

Abstract: Recently, the need of improved resource trading has arisen due to resource limitations and energy optimization problems. Various platforms supporting resource exchange and waste reuse in industrial symbiotic networks are being developed. However, the actors participating in these networks still mainly act based on predefined patterns, without taking the possible alternatives into account, usually due to the difficulty of properly evaluating them. Therefore, incorporating intelligence into the platforms that these networks use, supporting the involved actors to automatically find resources able to cover their needs, is still of high importance both for the companies and the whole ecosystem. In this work, we present a hybrid recommender system to support users in properly identifying the symbiotic relationships that might provide them an improved performance. This recommender combines a graph-based model for resource similarities, while it follows the basic case-based reasoning processes to generate resource recommendations. Several criteria, apart from resource similarity, are taken into account to generate, each time, the list of the most suitable solutions. As highlighted through a use case scenario, the proposed system could play a key role in the emerging industrial symbiotic platforms, as the majority of them still do not incorporate automatic decision support mechanisms.

Keywords: hybrid recommender systems; industrial symbiotic networks; case-based reasoning; waste optimization; energy consumption optimization

1. Introduction

Intelligent Decision Support Systems (IDSS), and especially Recommender Systems (RSs), are being used in the majority of current application domains, where users come across a lot of alternatives and information that have to discover, process, and use. RSs have been identified by Ricci et al. [1] as among the most promising techniques that are able to handle the current information overload and support users in their decision making processes. In recent years, these techniques have gained ground and their use has been extended to various application domains facing information overload problems.

Industrial Symbiotic Networks (ISN) form an example of such domain, as the actors participating in these networks usually act based on predefined patterns due to the difficulty of properly identifying all possible symbiotic alternatives. Industrial symbiosis (IS) has been defined by Chertow [2], as "engaging traditionally separate industries in a collective approach to competitive advantage involving physical exchange of materials, energy, water, and by-products". Furthermore, ISN have proven to be

successful in waste treatment and re-use as well as in creating new business opportunities. Therefore, the recommendation of effective possible synergies, being "matches between companies interested in providing and using their waste", that the users were not aware of, would lead to both economic and environmental benefits for the companies and the sector as a whole. Therefore, the ability of a platform to deliver such recommendations is crucial, as companies many times are not willing to change their established workflows if there are not sure about the outcome.

This work forms the extension of our previous work presented in [3]. In this paper, we describe, in detail, our hybrid RS [4] initially designed and developed for the needs of the SHAREBOX project. The aim of this system is to support industrial users in identifying possible resources and symbiotic partners of which they were not aware, and may provide them with improved performance. The implemented recommender first uses a graph-based similarity model to identify resource similarities that are based on their European Waste Codes (EWC). Subsequently, it performs a hybrid case-based recommendation to find the solutions best matching the user queries. Various criteria, like the user resource needs, the current situation of the system, as well as proper resource characteristics, are taken into account in order to generate the list of the most suitable solutions for a user at a given moment. Although it has been implemented for a specific platform, this system could be easily adapted to be used as a complementary tool in similar Industrial Symbiosis IDSS.

The structure of the paper is as follows: in the next section, the theoretical background related to energy efficiency and Industrial Symbiotic Networks, as well as Recommenders Systems and Case-based Reasoning, being the core methodology used in our recommender, can be found. Following, in Section 3, the developed system is presented and, in Section 4, a case study highlighting its use in an Industrial Ecosystem can be found. Finally, in Section 5, we discuss the results presented and their importance towards energy consumption optimization, while in Section 6, we conclude and present the basic points of our future work.

2. Background

2.1. Industrial Symbiosis and Sustainability

During the 20th century, the average temperature of the earth has increased by 0.6 degrees, and it is estimated that it will increase 1–5 more degrees in the next century [5]. Green House Gas emissions (GHG) and solid waste management are major causes of it and they require urgent measures to reduce current trends [6]. The first one, mainly related with fossil fuels combustion, is related with both transportation and industrial emissions [5]. Apart from the obvious environmental impact of human activities, air and soil pollutants have, in turn, impact in health [6]. The need to find sustainable resources that assist initiatives to reduce global warming and protect the environment, and indirectly human health, is paramount. The change of paradigm in energy towards renewable sources [7] and the reduction of waste emissions and primary resource use in resource intensive industries is suggested as one of the critical pathways to accelerate sustainable development [8].

Process integration emerged as a powerful engineering-oriented methodology in the 1970s to achieve substantial reductions in industrial energy, water, and utility use [9], and led to numerous design methods and industrial applications [10]. PI is a holistic approach for engineering design that aims to minimize resource and energy consumption. It takes a bottom-up perspective that is based on thermodynamics. The core principle of PI is to consider a system holistically; individual components are integrated to achieve synergy for the benefit of the whole system [11]. In [12], an optimization tool that is based on evolutionary algorithms is proposed to optimize the process gas network in an integrated steel plant, where cost and CO_2 emissions are simultaneously minimized. In [13], a process is proposed to generate low cost activated carbon from natural resources, like nuts shells, to eliminate heavy metals from industrial wastewater.

In Europe, some directives promote the use of biofuels [14] and renewable energies [15] to reduce GHG emissions. In [6], recycling was proposed as one the main ways of reducing waste. Net GHG

savings from paper recycling and composting were quantified between 50 and 280 Kg CO_2 eq/ton MSW. For other materials (glass, plastic, ferrous metals, textile, and aluminum), recycling offers an overall GHG flux saving between 30 and 95 Kg CO_2 eq/ton MSW [16].

European Council Resolution of 24 February 1997 on a Community strategy for waste management confirms that waste prevention should be the first priority of waste management, and re-use and material recycling should be preferred to energy recovery from waste, where and insofar as they are the best ecological options. A number of European Directives, resolutions, and decisions of European Council between 1999 and 2006 establish legislation and recommendations for waste management and for reducing waste emissions at specific aspects, like periods of waste storage before disposal to reduce waste going to landfills [17], end of life vehicles [18], electrical equipment [19], batteries [20], etc. Directive 2006/12/EC on waste provides a general framework that requires waste management plans for each establishment.

European Directive 2008/98/EC on waste is an improvement of D/2006/12/EC were protection of environment and human health comes into the priority and moves from waste management to reduction of waste emissions encouraging member states to promote re-use and recycling of industrial by products for sustainability reasons. In this context, circular economy and circular ecology, in particular, becomes a priority.

The term circular economy was introduced for the first time in an occidental context in 1980 [21] to describe interactions between the economy and environment in a closed framework. In Germany and Japan, circular economy is based on the management of waste through the 3R (reduce, recycle, reuse), transforming the traditional productive cycle (resource-product-waste) into a circular flow (resource-product-recycled resource), thus reducing resources and waste, in line with industrial ecology field [22]. In 2012, the Ellen MacArthur Foundation report highlighted the economic and business opportunity of a circular restorative model and detailed the significant potential benefits for the European Union announcing cost savings of net materials, by 2025, worth up to \$630 billion [23]. In 2015, EC adopted a Circular Economy Action Plan (COM 2015 (614)) [24] promoting circularity as a development backbone for a sustainable Europe. Very recently, in March 2019, the 54 actions under the action plan have been completed or are being implemented [25]. The plan establishes a concrete and ambitious program of action, with measures that cover the whole cycle: from production and consumption to waste management and the market for secondary raw materials and a revised legislative proposal on waste.

Circular economy is interested in creating a circular flow of materials for a more efficient economy with fewer pollutants and other unwanted by-products, and industrial ecology, as popularized in [22], is the part of circular economy focusing on the waste and by-products management recycling and reuse in the industry. More specifically, industrial Symbiosis [26] is a subset of industrial ecology that is based on the idea of sharing mutually profitable transactions, where the by-products of some processes can substitute resources in other processes, and constitutes an excellent opportunity to contribute to the European Circular Economy Action Plan. Industrial symbiosis targets the industrial ecology goals through actions between firms [27]. In 2012, [23] reviews the original concept from Chertow [27] by eliminating the requirement of geographical proximity between symbiotic partners and including the possibility of involving non-industrial partners into the loop. The work in [28] provides some guidelines for policy development that promotes industrial symbiosis. The European Union highlights Industrial symbiosis as a methodology that stimulates industries to become more sustainable and recommends the development of platforms that helps industries to identify possible successful synergies. In [29], a recent survey of information systems assisting in the identification of industrial symbiosis is presented. In [30], the impact of knowledge injection into the industrial symbiosis assistants is explored and the interest of using Artificial Intelligence, electronic institutions, and semantics [31] to discover new synergies and promote eco-innovation is discussed. The SHAREBOX project (http://sharebox-project.eu/) proposes an integrated intelligent assistant for industrial symbiosis, which is well aligned with the need of developing stakeholders platform aimed in EC circular economy action plan.

2.2. Recommender Systems

Recommender Systems (RSs) are tools and techniques for information retrieval and filtering, used to suggest items to be used or consumed by a user. They interact with the user to provide meaningful, effective, and personalized recommendations of items that might be of interest to the user. RSs either generate a set of personalized recommendations/suggestions of items that are expected to be useful for a certain user or try to predict whether a specific item will be of interest to a user or not, based on his/her previous preferences and those that are observed on similar users. In their simplest form, the recommendation is provided in form of a list of ranked items. The term "item" refers to the entity that was recommended by each recommender (ex: products, songs, web pages, services, etc.), while user refers to the entity (a person or another system), which interacts with the RS through some specific interface, in order to receive the recommendations. RSs help users with no or limited personal experience and knowledge in a specific area and, therefore, provide recommendations for users with a lack of ability to evaluate or select among the offered items [1,32,33].

The rapidly evolving functionalities of the Web, along with the wide use of Internet and other networking services in recent years, enabled sharing an increasing amount of information about many items of various types. The amount of information available, and the easy possibilities of collecting these kind of information and access to it, enabled users in the advanced search and review of item characteristics. On the other hand, the resulting information overload has increased the complexity of encountering and properly handling the required and correct information. Thus, the need of developing an intelligent RS, which is able to effectively support users in handling the information overload and decreasing the complexity of decision-making is denoted in many areas. Through the proper data selection and analysis, RSs support users' decision-making processes, augmenting their ability as well as the quality of their decisions by enabling them to find items that presumably they will like to use [34].

The main components/data important for the use of a RS, can be divided into [32]:

- Background data: information/data required by the system before the instantiation of the recommendation process (may be related to users, items context as well as to the recommendation process itself).
- Input data that is the information provided by the user to the system to receive recommendations.
- The core recommendation component, the algorithm used that combines and processes the background and input data to produce meaningful recommendations aligned with the active user's request. The functionality of this component heavily depends on the type and scope of the RSs, as well as the data that this system is asked to process.

The problem to be solved, or the central development idea to be supported by a RS is that a user wants to find the items that are able to best support his needs and their use will maximize the utility that was observed by the user. Based on the hypothesis of user rationality, a rational user who is aware of the alternative item choices would always select the item/situation that maximizes his/her utility under certain circumstances. Therefore, the intention of a RS is to estimate the utility values of different items' use through the scope of a specific user and suggest him/her the item(s) that are most likely to provide maximal utility, therefore being preferred by him/her [34].

The most widely used recommendation methodologies in commercial applications can be mainly divided into *Collaborative Filtering* techniques and *Content-Based* techniques.

Collaborative Filtering (CF) is based on recommending the items among those liked by users ("neighbors") similar to the active user. CF techniques have been identified among the most successful approaches for RSs and they have been successfully used in data mining cases as well as in e-commerce applications. Based on the hypothesis that users have a stable buying behavior along time, i.e., when two users rated some items similarly in the past, they will probably evaluate other items in the future in a similar way as well. User profiles' are built of the items that have been highly rated by them, and the similarity of users' tastes is deducted from their previous ratings. For a given request, these

techniques first identify the neighbors of the active user, based on their previous ratings, and then from the set of their highly rated items recommend those that the active user has not yet tried. Although widely used in commercial applications, collaborative RSs still have to overcome the scalability and cold-start problems that limit their performance [34,35].

On the other hand, *Content-Based (CB)* techniques recommend to users items that are similar to those that they have shown preference in the past. In these systems user profiles are built from the characteristics of the items that a user has rated highly, and the items that she/he has not tried yet are compared to them. The items with the higher estimated possibility of being liked by the active user, as deducted from his/her past preferences, are then recommended. As CB techniques rely on more specific information about users and items, the creation of appropriate user and item profiles becomes more crucial. A profile that accurately reflects user preferences increases the system's recommendation effectiveness. This is something that has become of great importance in the recently evolved business strategies in e-commerce. These techniques can recommend new items that could cover users' needs. However, as they always recommend items that are similar to those that a user has already used and liked, recommendations' can be limited on diversity and they can be overspecialized as well [33,34].

Other recommendation techniques, like Knowledge-based (Case-based, Constraint-based), Utility-based, Rule-based, Demographic, Context-aware, semantically enhanced, and especially hybrid, being combinations of different techniques, are also considered, due to the extended use of RSs in various areas, along with the shortcomings that the two previously described methodologies come with [4].

2.3. Case-Based Reasoning

Case-Based Reasoning (CBR) is a problem solving paradigm that uses old experiences to solve new problems. It is based on the following sentence, also known as the CBR assumption, "Similar problems have similar solutions". A Case-based reasoner solves new problems by adapting solutions that have been successfully used in the past to the new situation [36].

It is a computational metaphor of the reasoning by analogy process, so frequently used in human reasoning. Indeed, CBR is closely related to the human way of thinking, reasoning, and acting through everyday situations when facing new problems that have to be solved. In such cases, there may be observed the tendency of recalling similar cases that have been successfully solved in the past or to avoid repeating the errors performed. Therefore, CBR, as a method for building intelligent reasoning systems, seems to be more natural and constitutes a problem solving methodology that can be applied to many domains, as it does not require building an explicit domain model. In addition, as CBR reasoners derive their reasoning from complete cases, rather than decomposing them into rules, they have the ability to adapt and improve their problem solving performance along time, in comparison to rule-based techniques, that contain dependencies that cannot be easily understood, and may turn to be insufficient for some domains of applications, without the use of expert domain knowledge [37].

In CBR, a case denotes a problem situation in a wider term; it may be any problem that is defined by a user that does not necessarily refer to finding a concrete solution to an application. A case can be defined as a set of values of specific characteristics that occurred in a situation. Depending on the type and the characteristics of the situation that is modeled, a different case representation is used to capture these specific attributes that specify this case and their values. Often, cases may be described and stored as collections of attribute-value pairs that can be easily stored and retrieved through the CBR cycle. However, in more complex situations, it is useful to use a more complex description, like a hierarchical object-oriented representation of the cases, where cases are represented as a collection of structured objects, instances of a class, which enables their decomposition and analysis as well as the use of inheritance and the extraction of possible relations among the objects parts. In such an object-oriented approach each object represents a closed part of the situation, each belonging to a class and being described by a set of features. Finally, for special applications, a graph representation might

be adequate, where a case is represented as a set of nodes and arcs or in others predicate logic may be used in order to represent the cases as sets of atomic formulas [38,39].

The selection of the most appropriate cases' representation in a problem depends on the problem domain and on the scope of the CBR system, as well as on the amount and structure of the already available data that form the case base. Depending on the complexity of the representation used, the complexity of the used similarity metric will also vary.

2.3.1. CBR Cycle

The CBR solving and learning process can be described as a cyclical process comprising from 4 parts. The following four processes, also known as "the four REs" (Retrieve, Reuse, Revise, Retain), or as the CBR cycle are [36]:

- *Retrieve:* the most relevant cases among those previously experienced from the case memory.
- *Reuse (or adapt):* the information and knowledge provided by the by the retrieved case(s) in order to solve the new problem.
- *Revise (or evaluate):* the solution obtained
- *Retain (or learn):* the parts of the solution/experience that are likely to be used (reused or avoided) for future purposes and incorporate this new knowledge into the case base.

The CBR cycle can also be seen in Figure 1, below, where as we can see except from the knowledge obtained by the cases in the case base there is also general, domain dependent knowledge present, supporting the CBR process. Each of the processes in the cycle is further divided into subtasks, depending on the type of the application domain.
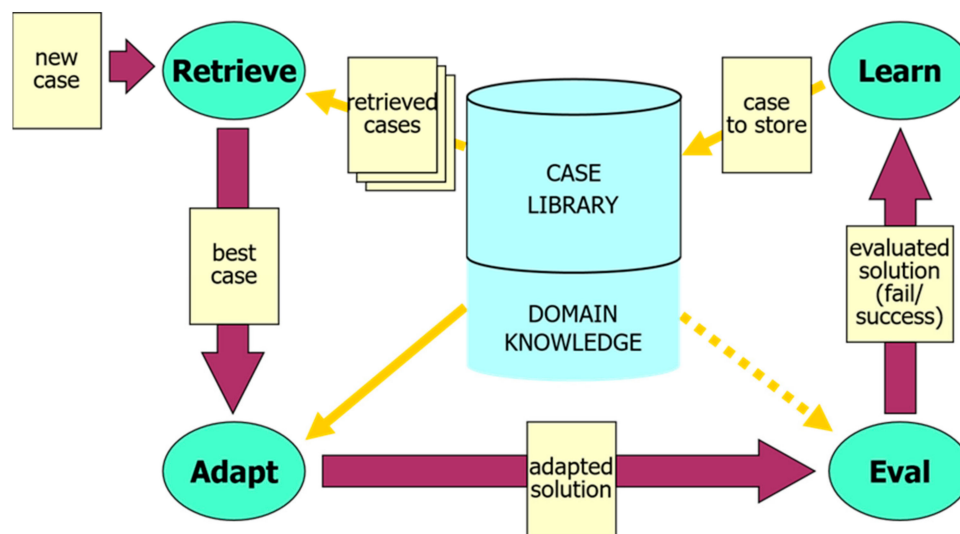

Figure 1. The Case-Based Reasoning (CBR) cycle.

When a new problem ("new case") asking for solution comes, the CBR approach begins with retrieving one or more of the previously experienced similar cases among those that were recorded in the case base. This step also requires an indexing of the existing cases based on appropriate features, therefore similarity measures are involved in this step. The solution is obtained by reusing the most similar among the previous cases that were retrieved after adapting it, so that it becomes adequate for the new problem. The solution can either be modified manually by the users wishing to define some characteristics of interest to them, or automatically by the system based on domain knowledge and solution generators that are able to adapt the solution to the special requests. After being derived, the new problem solution is evaluated in order to ensure that it is adequate for the initial problem and to verify that its quality and performance will be the expected. Finally, the new experience is added into

the existing case base by providing additional solution knowledge or an explanation of a way that performed mistakes can be avoided in the future [39].

2.3.2. The Similarity Concept in CBR

A core concept of the CBR methodology and a key factor of its successful application is the similarity measure that is used for cases retrieval, i.e., the measure used to quantify the degree of resemblance between a pair of cases. The purpose of the use of similarity metric is to select those cases that are most similar to the new case from the case base. Therefore, they may have the same solution as the current problem or their solution can be easily adapted to match the characteristics of current problem. The similarity metric provides an apriori approximation of the rate of the utility the solution is going to provide to its reusability, with the intention to provide a good approximation that is as close to the real value of reusability as possible, while at the same moment being easily computable and interpretable. Cases can be represented in different ways, and several similarity metrics can be used for each case representation. The values of the similarity function normally range in the interval [0, 1], with 0 being assigned to totally different cases and 1 to cases that are regarded as identical through a concrete similarity measure [38]. Any normalized metrics adapted to the nature of the cases are suitable for a CBR system.

There have been various similarity metrics proposed in literature [40]. Usually, the notion of a metric or distance $d(x, y)$ between two objects, x and y, has been used in order to reflect the level of similarity or dissimilarity among the elements in a given set, where, through the relation $S(x, y) = 1 - d(x, y)$, distances $d(x, y)$ can be transformed into proper similarity measures $S(x, y)$, provided that $d(x, y)$ evaluates in the interval [0,1]. In order to build the CBR model, there is a need of retrieving the k most similar cases. The k-nearest neighbors' algorithm approach can be used to specify these cases. When considering a set of n features f_i, $i = 1, \ldots, n$ used to describe the objects and evaluating in the space [0, 1] and being w_i the importance weighting factor of a feature f_i, the similarity between the input case (represented by $(f_1^I, \ldots, f_n^I)$) and a retrieved case from the case base (represented by $(f_1^R, \ldots, f_n^R)$) can be represented as the weighted average of the local similarities feature by feature $(sim(f_i^I, f_i^R)$:

$$\frac{\sum_{i=1}^n w_i \times sim(f_i^I, f_i^R)}{\sum_{i=1}^n w_i} \tag{1}$$

Based on this, most case-based reasoners use a generalized weighted dissimilarity measure that can be defined as [41]:

$$d(f^I, f^R) = \frac{\sum_{i=1}^n w_i \times (f_i^I - d(f_i^I, f_i^R))}{\sum_{i=1}^n w_i} \tag{2}$$

where $sim(f_i^I, f_i^R) = (f_i^I - d(f_i^I, f_i^R))$ and $d(f_i^I, f_i^R)$ describe the dissimilarity level between the i-th feature of the compared cases.

In cases that are characterized by symbolic or Boolean attributes, or by attributes of more complex types, different and often more complicated local similarity functions have to be defined, more than the distance measures. Commonly, the case description contains both numerical and qualitative information. In these cases, compatibility measures are appropriated [42]. When a semantic interpretation of the qualitative terms is also available, through some formal representation, like a reference ontology, semantic distances can be also used [43]. Additionally, the qualitative terms can be used as tags and co-occurrences can be exploited for the computation of case similarities.

2.4. Case-Based Recommenders

Case-Based Recommenders have mainly emerged as an alternative to CF recommenders intending to overcome the shortcomings that CF recommenders come with, while efficiently handling the existing information overload problem.

The predominantly used CF recommendation techniques focus on users' preferences, as these can be expressed through users' purchase histories and item assigned ratings, without taking into account the general context in the moment of the item selection or the market situation at the transaction moment. The recommendations are derived from the items that users, similar to the active user, have selected in the past and evaluated positively, more than identifying and analyzing the attributes of the available items. Therefore, item ratings more than item attributes' descriptions are required by this type of recommenders.

CBR recommenders, on the other hand, will generate recommendations for an active user that are based on the analysis of the item characteristics and by trying to find the item(s) that best match a user request. These recommenders include in the recommendation generation process semantic ratings and characteristics, as these can be extracted from past item selection cases with similar user requirements, placing their emphasis on the description of the requirements and the characteristics of the cases. CBR recommendations are able to provide accurate results when applied to domains where the individual products are described in terms of a well-defined set of features [38].

Case-based recommenders implement a type of content-based recommendation that relies on a structured representation of items as a set of well-defined characteristics/features and their values, in contrast to general content-based recommenders that usually rely on unstructured or semi-structured items' representations. These representations allow for case-based reasoners to make judgments about product similarities and based on those to provide recommendations, more than simply using the ratings that are assigned to products by various users. The existence of a structured and common way of representation of the treated items enables case-based recommenders in calculating and understanding the similarities among those items, the generation of meaningful item recommendations of high quality to the users while enabling the evaluation of the outcome in terms of user satisfaction and the incorporation of this feedback.

Case-based recommenders rely on the core concepts of retrieval and similarity of CBR. For CBR recommender, the user query serves as a problem specification, while the item descriptions form the cases in the case base. The set of existing items is represented as cases of which the case base is made up, while the item(s) to be recommended to users are the items (cases) that are retrieved form the case base, based on their similarity to the user's request as this is defined in the same space, the space of items' characteristics. These characteristics depend on the type of the specific item traded in each situation. Depending on the type of the item to be recommended and its special characteristics, different local similarity measures are aggregated and modeled into a different global similarity function that is then used for the final item selection and highly affects the outcome of this process [44].

3. System Description

3.1. The SHAREBOX Platform

The main functionality of the SHAREBOX system is to allow a user, which usually will be an employee of one company, to specify the resources he/she WANTS or HAS, with the final aim of obtaining some possible synergies with other companies.

A *synergy* is a symbiotic relationship between two companies where one resource or waste produced by one company (i.e., a HAVE of the company) can be consumed by the other company, which needs it for productive activity (i.e., a WANT of the other company). The SHAREBOX platform lets the user introduce the characteristics of the resources WANTED/HAD, which are usually defined by some descriptive data, like EWCs codes, some keywords, etc., and for each selected resource WANTED/HAD, the user asks to the system for possible synergies with other companies.

Currently, there are several modules implementing different approaches to calculate the suggested synergies. Some of these components, especially some input-out modelling techniques, could be very hard from a computational point of view, particularly when the resource database of the SHAREBOX platform will be highly populated. This seems to leave space for trying to *improve the computational*

time spent to obtain the suggestion of the synergies for a given company. On the other hand, there is another important feature that can be improved: *the quality of the recommended synergies*, because an intelligent component could learn from the past synergies performed through the platform system. This way, each time that a synergy is finally completed, a new valuable experience and information can be learnt by this intelligent module, issuing at improving future recommendations, and also learning which synergies were not successful, aiming at not making the similar errors in the future. Currently, the SHAREBOX architecture can be outlined, as depicted in Figure 2.
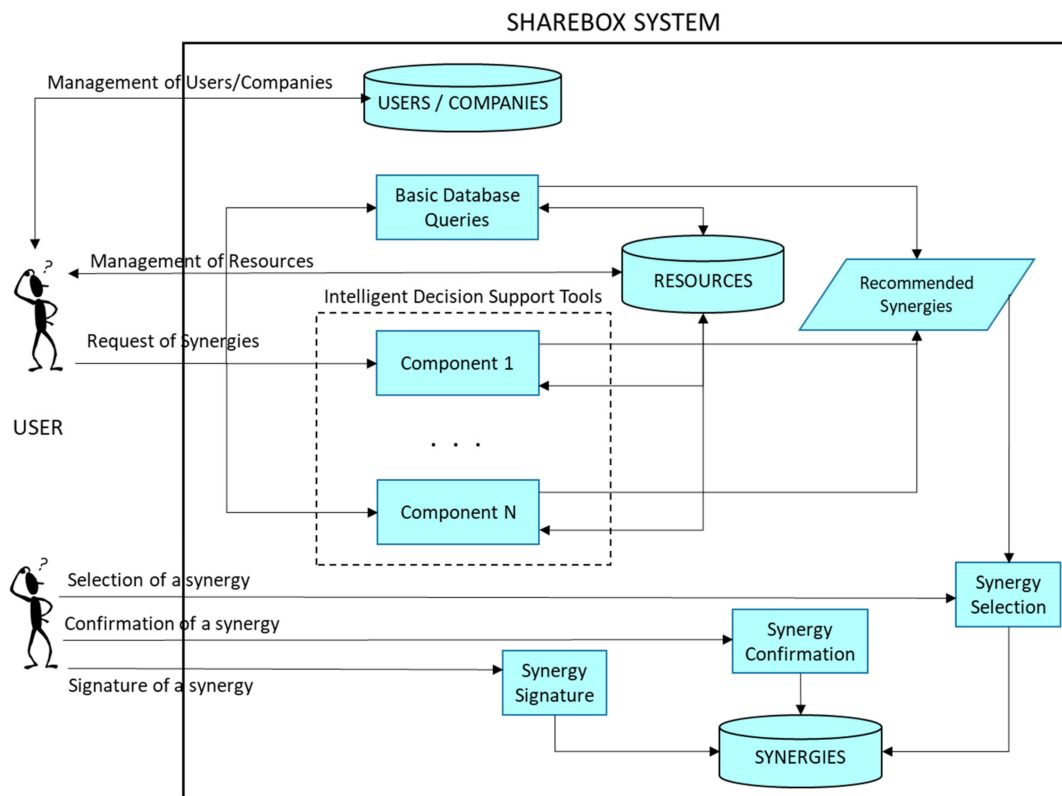

Figure 2. Current SHAREBOX System Architecture.

3.2. Hybrid Recommender System Overview

In Figure 3, the new version of the platform including the recommender is shown. Although the system was designed for the SHAREBOX platform it could be easily adapted to other platforms serving industrial symbiotic networks with similar needs.

3.3. Scope

The scope of the presented recommender system is to generate the list of top-N recommendations in terms of possible synergies matching a given user request related to a specific resource and the current resource availability status. More specifically:

- Given:

  ○ A list of company sites, with their industry codes, geographical locations, resource needs, and offers
  ○ A query that consists of, at least, the site location, the resource type, in terms of European Waste Code (EWC), category and the definition of whether it is a needed or offered waste

- Generate the list of the N most adequate resource synergy candidates, those from the available matching in a locally optimal way the given query.

3.4. Problem Data Entities

In general, we refer as a "user" to the entity that interacts with the recommender that might represent a company site or a company. The main problem entities are the companies' sites and the resources registered by these sites that are presented in more detail in this section.

Figure 3. Overview of the SHAREBOX Platform with the Recommender.

3.4.1. Current System Status

The recommender first has to analyze the current situation of the system in order to generate recommendations for a new query. To this direction, the list of the currently available resources in the system is loaded. Every record in this list refers to a resource offered or needed in the given moment. Therefore, this list is updated every time a new resource need is registered in the system.

The important data contained in this list can be described by the following attributes that are then used to map towards the given query. This data records contain both company and resource related information that is taken into account during different phases of the recommendations generation.

Resource related:

- *Name*: identifier of the resource
- *Description:* basic information describing the resource
- *Keywords:* automatically and user assigned
- *Waste code(s):* EWC(s) of the registered resource
- *Direction (Have/Want):* defines whether the resource is offered or needed by the company site
- *Quantity:* maybe a specific number or in some cases it is possible to have no limitations in the resource production, in terms of theoretically unlimited availability

  ○ The quantity may also be defined after two parties have expressed mutual interest in forming a synergy and negotiations have been initialized

- *Measure unit:* depending on the type of resource

- *Availability:* refers to the frequency by each it is generated or requested. Mainly, there are resources of continuous availability, therefore from the moment these are registered they are always generated/needed, thus in any time moment may be provided or are always needed
- *Site:* each resource, rather than to a company, is associated with the site where it can be found

Site related: this information is used after the candidates identification process in order to enable the users communication in case that one(s) of the returned resource recommendation seems as suitable and the user wants to obtain more information regarding the providing or requesting company and possibly establish synergy.

- *Industrial code:* defining the industrial category of the site
- *Resources:* that has previously registered on the platform as needed or offered
- *Address:* the geographical location of the site

  ○ Presented by the geographical coordinates, longitude and latitude

- *Company:* that this site belongs to

3.4.2. User Query

The platform user usually represents a company site and has been previously registered into the platform, and accesses the platform in order to find a specific resource that would need to use or to find possible ways to distribute its available waste.

The following information is used to build each time the user query in order to provide support to the users and identify the resources being most adequate for them on a given moment:

- *Keyword(s):* One or more keywords describing the resource requested
- *Direction:* Refers to the type of need, whether it is about a needed (want) or an offered (have) resource
- *EWC:* The waste code of the requested resource

  ○ Given in the three hierarchical levels, the chapter, the subchapter and the code.
  ○ Search method (Filter or Rank): Refers to the extraction of resource candidates, whether only those with EWC exactly matching the one in the query can be used (filter), or whether the similarity of the EWCs of the existing and the requested resources will be evaluated (rank). In the second case, the resources are ranked based on advanced similarity functions applied to the EWCs of the resources available and the requested one.

- *Category:* The resource category of the requested resource

  ○ Presented in three hierarchical levels
  ○ Search method (Filter or Rank): Same as for EWC, but evaluating the exact match or similarity of the resources' category with the one in the query.

- *Site:* the geolocation of the site where the resource is located or needed. The site does not refer to the company, but to the exact resource, as companies might be registered by their head offices, while they may have various sites where their resources are stored or various processes where the resources are needed.

  ○ Presented by the geographic coordinates, longitude and latitude.

3.5. Query Resolution

In the proposed model, the following relationships are used in order to approximate the degree of similarity of resources, to identify the possible resource alternative candidates and to propose matchings.

- EWCs are associated with keywords that describe them.

- Resources are associated with their attributes where the EWCs are one of these attributes.
- Company sites are associated with the resources they have and could offer or would need, with a given frequency or at a specific time moment.

3.5.1. Search Method: Filtering or Ranking

According to the search method that was selected by the user, which might refer to a filtering or a ranking process, the recommendation algorithm may return a different set of results. In case a user has selected results filtering, simply the resource candidates that have been identified in the submitted request that could be of interest to him/her are presented. On the other hand, when a ranking process is required, an ordered list of the top-N resource candidates is generated while more parameters taking into account, like the similarity of the EWCs of the resources, the distance of their sites, etc.

3.5.2. EWC Analysis and Similarities—EWC Representation and Similarity Functions

EWCs use six digits to describe waste based on the processes from which they have been originated, through three levels (two digits for each) that finally lead to:

- 20 Chapters—Business sectors
- Subchapters—Processes
- Codes—Waste descriptions

Example: 10 01 26

In the majority of the systems, based on the EWC(s) submitted by the user, only the exact matches (if any) would be presented as possible solutions. Thus, an important functionality of the designed system, which is able to deliver an added value, is its ability to identify alternative resource candidates that may be further evaluated based on their specific characteristics. As the EWC categorization is based on the processes and the industrial sectors from which waste comes, and not on its recourse properties, waste with identical properties may be also classified under different codes.

The resources that were registered in the system are associated with various characteristics. Among the most important is the resource EWC, as it is the attribute that mainly defines the degree of suitability of a resource for a given process, while the other characteristics of a resource are more related to its availability in economical and temporal terms.

This model follows the graph model used in [45,46] to identify strong similarity that is based on the tags assigned to songs, being metadata, sound properties, or user tags. Here, we use the keywords as tags that are assigned to EWCs that are further associated with resources. Each EWC is associated with keywords coming from the Sepa resource Thesaurus [47], a proper dictionary, thus "tagging" codes with these keywords does not insert additional noise into the model, as usually happens with user generated tag clouds. Therefore, a graph-based model was used to describe first the EWCs through the keywords of the existing thesaurus with which each EWC is mapped, like in Figure 4, below, where the keywords in common have been also highlighted.
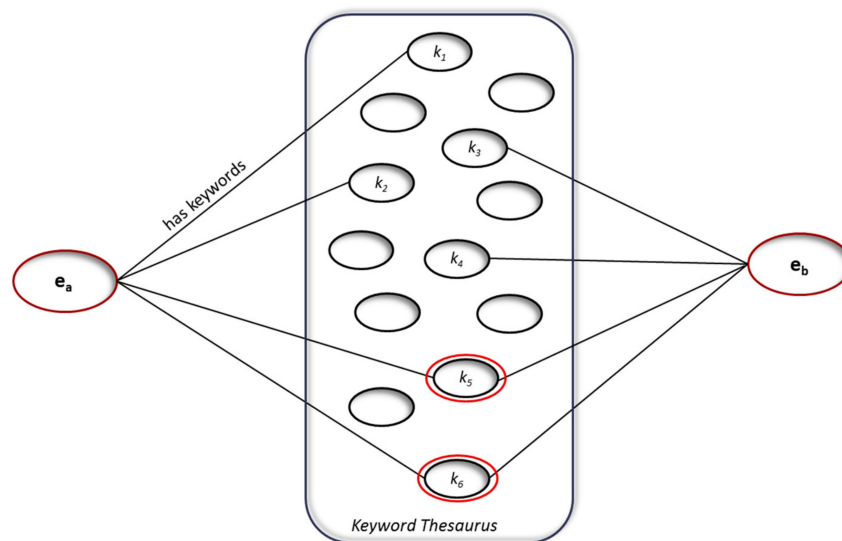
Figure 4. European Waste Codes (EWCs) graph based representation using keywords from the associated Thesaurus.

The similarity of two codes is calculated as a function of both the density of the common tags in their descriptions and the unique terms of each of them. This modelling permits us to identify the clusters of item codes with common descriptions that could be automatically recommended to serve the same request, as well as to identify groups of similar items, which share some part of their descriptions, which could possibly be used, if a user decides so.

Let two EWCs be e_a and e_b having $n(a \cap b)$ tags in common, while $n(a \backslash b)$ and $n(b \backslash a)$ being the number of those that are only associated with e_a and e_b, respectively. Their similarity is calculated based on the density of their common descriptions, as in [45], while using Equation (3):

$$sim_{Gb}(e_a, e_b) = 1 - log_2\left(1 + \frac{n(a \backslash b) + n(b \backslash a)}{n(a \cup b)}\right) \tag{3}$$

directly related with the mutual information between EWCs. The two EWCs have no tags in common, $n(a \backslash b) + n(b \backslash a) = n(a \cup b)$ and the similarity becomes $1 - log_2(1 + 1) = 0$. On the contrary, when all of the tags are in common $n(a \backslash b) = n(b \backslash a) = 0$ and the similarity becomes $1 - log_2(1) = 1$.

Additionally, EWCs are organized according to a hierarchical structure that describes semantic relationships between them (see Figure 5) [47]. The first level of the hierarchy collects the two first digits of the EWC and it corresponds to chapters, as said before, which means the business sector (like 04: Wastes from the leather, fur and textile industries or 10: Waster from the thermal processes). The second level of the hierarchy represents digits 3 and 4 of the EWC and corresponds to the subchapters or to the description of the industrial process (Chapter 04 contains subchapters 01: Wastes from the leather and fur industries and 02: Wastes from the textile industries, whereas Chapter 10 contains 14 different subchapters like 02: wastes from the iron and steel industry or 11 wastes from manufacture of glass and glass products). In the third level, the description of the waste produced in each specific industrial process is described and this corresponds to digits 5 and 6 of the EWC (Thus, EWC 040210: organic matter from natural products (for example grease, wax), 040222: wastes from processed textile fibres, 100201: wastes from the processing of slag; 100211: wastes from cooling-water treatment containing oil, 101105: particulates and dust; 101115: solid wastes from flue-gas treatment containing dangerous substances). Resources, belonging to the same sector or coming from the same process also share some characteristics and properties. Therefore, having common sub-codes, as highlighted in Figure 5, is also a sign of resource similarity, even if the entire path from the root to the leaf is not the same. A new similarity function is defined to capture the impact of this hierarchy in the similarity.
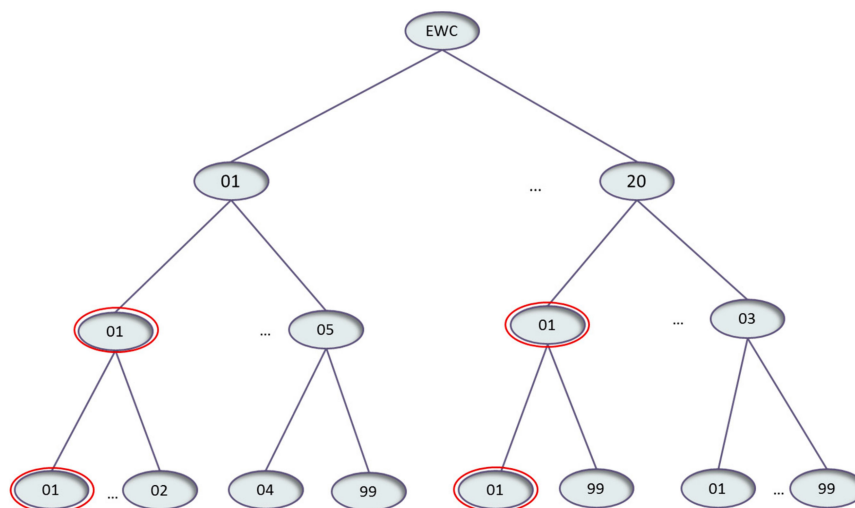
Figure 5. EWCs hierarchical representation.

Equation (4) has been used to capture this similarity notion:

$$sim_{Hb}(e_a, e_b) = \frac{1}{6} \times \sum_{l=\{S,P,W\}} (3 - p(l)) \times sim_l(e_a, e_b) \tag{4}$$

where *l* represents the three levels of the EWC ontology represented in Figure 5 and takes three values: *S* for the higher level indicating Bussiness Sector, *P* for the intermediate level indicating the industrial process and *W* for the third level indicating the waste. Function *p(l)* provides the position of the level ($p(S) = 2$, $p(P) = 1$, $p(W) = 0$), and $sim_l(e_a, e_b)$ refers to the similarity between the EWC in the digits corresponding to the level l. For l = S, the digits 1 and 2 of the EWC are considered, for l = P digits 1, 2, 3, and 4, and for l = W digits 1–6 are considered. The term $sim_l(e_a, e_b)$ evaluates to 1 when the corresponding digits of the two EWC are equal and to 0 otherwise. Accordingly, the term sim_l is only activated when the compared EWC are equal for the corresponding level l and gains weight in the similarity with the depth of the coincidences in the ontology. This means that $sim_W(e_a, e_b)$ impacts the final similarity, with weight 3 when it is activated, whereas $sim_S(e_a, e_b)$ only with weight 1, catching the hierarchical structure of the EWC ontology and representing the idea that the more nodes in the ontology are, in common, higher is the similarity. The final value is normalized to interval [0,1] by the term 1/6.

Finally, the third similarity function that was used combines the previous two with equal importance in order to capture all possible similarity dimensions, being:

$$ewcSim(e_a, e_b) = 0.5 \times (sim_{Gb}(e_a, e_b) + sim_{Hb}(e_a, e_b)) \tag{5}$$

The similarities between EWCs have been calculated as part of a pre-processing phase that has been performed offline, and are stored in the system in order to lower the computational effort and speed the response time of the system. Therefore, each time a new query is submitted to the system, based on the required EWC the most similar EWCs are easily identified and then also based on those, the set of resource candidates can be found.

3.5.3. Category Similarity

Except from the EWC in the used platform, an additional parameter enabling resource classification is their resource category. The resource categories used also consist of three hierarchically structured levels. An expression that is similar to Equation (4) is used to calculate the similarities of resource categories, denoted as catSim.

3.5.4. Keyword Similarity

Registered resource records contain the resource name, keywords automatically assigned to the resource as well as (possibly) keywords, and a description added by the user that registered them. In general, resources might have more or less accurate descriptions. We first evaluate whether the keyword is contained in the resource name in order to approximate the degree of similarity of the keyword in the query and the linguistic information related to existing resources. Subsequently, in the case of additional appearances of the keyword, we evaluate the percentage of its appearance in the whole textual description, denoted as keySim.

3.5.5. Advanced Ordering

In the case that a ranking of the results is asked, a multi-attribute similarity function that evaluates the level similarity of the candidate resources previously retrieved with the user case, while taking into account the EWCs', the categories', and the descriptions' degree of similarity and also additional dimensions of the treated resources, like the distances of their sites, form the inputs of the similarity function used to identify the top candidates that will be presented to the user. As one or more fields in the query might be left blank, although we consider as being more informative the EWC similarity by placing a higher weight to it, we evaluate all the above described similarity dimensions in order to satisfy the user combined with the distance between the requesting site and the registered sites.

The rating values are finally normalized to take values in the interval [0,1], assigning 1 to the most adequate candidates.

3.6. Recommendations' Generation

3.6.1. Process Overview

An overview of the basic steps followed by the designed hybrid recommender system can be found in Algorithm 1 below, and they are presented in more detail in the next section.

Algorithm 1. Recommendations' generation algorithm.

Input:	Current resources' status L_R, new query q_N
Output:	TopN resource recommendations L_{rec}
1:	Initialize recommender: Load parameters
2:	Analyze q_N
3:	Retrieve $L_{FR} \subseteq L_R$: filtered resources
4:	**if** ($|L_{FR}| > 0$)
5:	**for** ($r_R \in L_{FR}$)
6:	**if** ($direction(q_N) * direction(r_R) \leq 0$)
7:	$Sim(q_N, r_R) = ewcSim(q_N, r_R) * w_e + catSim(q_N, r_R) * w_c + keywSim(q_N, r_R) * w_w$
8:	$rating(r_R) = Sim(q_N, r_R) + invDist(q_N, r_R) * w_d$
9:	Rank candidates in L_{FR}
10:	**if** ($|L_{FR}| > 20$)
11:	$L_{rec} \leftarrow top\_20(L_{FR})$
12:	**else**
13:	$L_{rec} \leftarrow L_{FR}$
14:	**return** L_{rec}

3.6.2. Resolution Workflow

For a given case described in terms of *a Resource r_i needed and/or a Waste w_j generated* at a specific company site, along with some characteristics of the resources and geographical data of the sites where these are located, in order to generate the list of the top–N matching candidates, the recommendations' generation workflow can be divided into the following basic steps:

1. A user inserts a *resource request*, containing the specifications of the desired resource, in terms of key words, EWC codes and resource categories, the resource direction specifying whether it is a desired or an offered resource, and finally the search method mode, defining it apart from the exact matching descriptions other similar resources can be considered as candidates. In addition, the site geolocation is retrieved from the stored records.

2. A first *filtering* of the resources is performed, based on the *request direction* parameter that has been specified in the user query. If the performed query is about a want waste only the records related to have waste(s) will be evaluated. On the other hand, in the case of a have request only the want records will be taken into account. If the user is simply browsing to see what resources are available in the system and no direction is stated, all of the resources are retrieved at this stage.

3. The recommender will then identify the *candidates* that could address the user request, in terms of resource properties. By candidates, we refer to the various alternative resources with characteristics that could serve a user request. First, if a user has specified a filtering mode at EWC or category level, first the available resources are filtered based on this parameter. Subsequently, this new set of resources is used for further evaluation.

4. Finally, the *core recommendation mechanism*, with scope to return the ordered list containing the closest matchings of the requested waste, is based on a similarity/distance function comparing the user's request(s) and the existing cases' parameters. Let *q* be the user request and *r* one of the existing resources after the filtering processes. The locations of the companies as these have been entered to the platform, the available and requested quantities along with their degree of similarity are among the parameters that are used in the distance function. More specifically, the total semantic distance among *q* and *r* will be a function of the following general form:

$$d(q,\ r) = f\big(sim\big(res_q, res_r\big), sim\big(loc_q, loc_r\big)\big) \tag{6}$$

where:

- $sim\big(res_q, res_r\big)$ is the similarity function that is used to calculate the semantic distance of the requested and the candidate resources, given that not only exact matches (the same resource) are considered as acceptable solutions, but also closely similar resources may be used. This function is a combination of the similarity functions (5) described previously and the terms described in Sections 3.5.3 and 3.5.4.

- $sim\big(loc_q, loc_r\big)$ is a function that evaluates the geographical distance between the requesting and the candidate sites' locations.

Usually, the resource candidates would be directly searched by the EWC(s) submitted by the user, and only exact matches (if any) would be presented as possible solutions. In order to provide additional possibilities to the users, this Recommender first identifies the associated, most similar with the requested resource code, additional codes. Under these codes items with same properties or closely similar can be found, although being classified under different codes. Due to the EWC categorization, which is based on the processes and the industrial sectors from which waste comes, and not on its recourse properties, many times waste with identical properties is classified under different codes. Thus, an important functionality of the system, which is able to deliver an added value to the platform's users, is the identification of those alternative resource candidates.

4. Case Study

In this section, the presentation of an industrial symbiotic ecosystem, being the application domain of our RS, can be found. In addition, there is the presentation of recommendations related to the resources that are involved in synergies with higher frequency in this ecosystem.

In industrial symbiotic networks there is a clear terminology regarding the different inputs needed and the different outputs generated by the industries. The *primary inputs* are those inputs that are

not produced by any industry, such as raw materials and energy. The *primary outputs* are the outputs resulting from a manufacturing process that is the primary focus of production. *Secondary output* generally refers to *waste* that has no perceived economic value and that traditionally are to be discarded. Some concepts are important, like the *substitution rate* of a primary input A (obtained from waste recycling) over primary input B, in a given production process, is referred as a/b if a units of A can be used instead of b units of B in the production process. The *recycling rate* of primary input C over waste D in a specific recycling process is referred as c/d if by means of recycling d units of D, we can obtain c units of C in the recycling process.

Usually, industries in a country are located in some industrial parks, where several industries are not far located in the same industrial zone, and usually they share some characteristic features, like, for instance, producing several products that can be used by a determined type of industry, like a ceramic industrial park, were several industries produce tiles, other produce some components that will be used by the tile industries as primary inputs, like some glass powder, others produce some mineral components, etc. Some real examples of industrial parks are:

- Industrial Cluster (La Mina Industrial Zone and Castelló Harbour) in Nules, Castelló de la Plana, Spain
- Industrie Center Obernburg, in nearby Frankfurt, Germany
- Uslan Eco-Industrial Park (EIP), South Korea
- TaigaNova Eco-Industrial Park in Fort McMurray, AB, Canada
- CleanTech Park, Singapore

4.1. Industrial Ecosystem

As a case study to analyze how the designed Recommender system works, an industrial ecosystem will be described. Let us suppose that we have a wide industrial park outside a great city with the following eight industries located in the park:

1. CeramProd, which is an industry producing ceramic manufacturing products, like Floor-Tile-Ceramics, Wall-Tile-Ceramics, and Roof-Tile-Ceramics, and it needs Frites, Recycled-Water, and Pigments.
2. NewFarm, which is a modern industry producing several outputs of a farm like Feathers, Leather-cuttings, Milk, and needs Wooden-Packaging and Glass-Bottles.
3. StoneRock, which is an industry producing several materials, like Chrome, Paint-Powders, Frites, and needs Mechanical-Equipment and Sand.
4. SweetHome, which is a modern industry producing several home items like Pillows, Bed-Sheets, Table-Clothes, and needs Feathers, Silk-Waste, and Wooden-Packaging.
5. HotelNew, which a hotel inside the industrial park which has Wooden-Packaging, produces Domestic-Wastewater and needs Pillows, Bed-Sheets, and Drinking-Water.
6. ConsMat, which is a modern industry producing materials for the construction, and mainly produces Bricks, and needs Sludge, Pigments, Paint-Powders, and Recycled-Water.
7. DrinkWater is a modern WasteWater Treament Plant including a tertiary treatment devoted to the production of DrinkingWater and also generates Sludge as a subproduct, and it needs Industrial-Wastewater, Domestic-Wastewater, and punctually it needs Steel-Pipes for some maintenance operations.
8. ChefRest, is a restaurant within the industrial park which produces Domestic-Wastewater, and needs Milk, Drinking-Water, Table-Clothes, and Textile-Packaging.

Their needs and outputs are described in Table 1.

Table 1. Description of the industries needs and output products within the industrial park.

Company	Product	H/W	Quant
CeramProd	Floor-Tile-Ceramics	Have	2000
CeramProd	Wall-Tile-Ceramics	Have	8000
CeramProd	Roof-Tile-Ceramics	Have	1000
CeramProd	Frites	Want	5700
CeramProd	Recycled-Water	Want	15,000
CeramProd	Pigments	Want	3750
NewFarm	Feathers	Have	2
NewFarm	Leather-cuttings	Have	50
NewFarm	Milk	Have	3000
NewFarm	Wooden-Packaging	Want	1500
NewFarm	Glass-Bottles	Want	2500
StoneRock	Chrome	Have	4150
StoneRock	Pigments	Have	6000
StoneRock	Paint-Powders	Have	1000
StoneRock	Mechanical-Equipment	Want	4000
StoneRock	Sand	Want	8000
StoneRock	Frites	Have	6100
SweetHome	Feathers	Want	5
SweetHome	Silk-Waste	Want	2500
SweetHome	Wooden-Packaging	Want	1000
SweetHome	Pillows	Have	200
SweetHome	Bed-Sheets	Have	400
SweetHome	Table-Clothes	Have	300
HotelNew	Pillows	Want	100
HotelNew	Bed-Sheets	Want	50
HotelNew	Wooden-Packaging	Have	1000
HotelNew	Cardboard-Packaging	Have	1000
HotelNew	Drinking-Water	Want	20,000
HotelNew	Domestic-Wastewater	Have	6000
ConsMat	Sludge	Want	20,000
ConsMat	Bricks	Have	250
ConsMat	Pigments	Want	3000
ConsMat	Paint-Powders	Want	250
ConsMat	Recycled-Water	Want	5000
DrinkWater	Drinking-Water	Have	25,000
DrinkWater	Industrial-Wastewater	Want	35,000
DrinkWater	Domestic-Wastewater	Want	4000
DrinkWater	Sludge	Have	8000
DrinkWater	Steel-Pipes	Want	6000
ChefRest	Milk	Want	1000

Table 1. *Cont.*

Company	Product	H/W	Quant
ChefRest	Drinking-Water	Want	5000
ChefRest	Domestic-Wastewater	Have	2500
ChefRest	Table-Clothes	Want	30
ChefRest	Plastic-Packaging	Want	2000
ChefRest	Wooden Pallets	Have	2000

Let us suppose that the company ChefRest, which has recently installed in the industrial park, currently needs 1100 L of Milk, 5200 L of Drinking-Water and 45 Table-Clothes. Usually, it had some supplier for the milk and the table clothes, which was very far from its location (approximately 200 km). In addition, the drinking water was supplied for a municipal company in its previous location, but the prix of water was expensive 5 €/L.

Of course, the idea, is that now, ChefRest, if they knew that in the same industrial park there are some other industries that produce the primary outputs they need, they will optimize the price they were paying, the transportation costs and the corresponding energy savings, and even they could take the Drinking-Water, from the nearby DrinkWater company, which is a Drinking Water Plant, which reduces the consumption of water and its equivalence in energy costs. All of these new resources, hidden until now for a user, can be made apparent to a user of the SHAREBOX platform, and the new wastes, which will be used as primary resources will be suggested by the recommender system integrated in the platform.

In the next section, the recommendations of our system will be detailed to outline the benefits of the industrial symbiotic networks, and the possible synergies that can be discovered or maintained in time to save costs, energy, and primary outputs through recycled waste materials.

4.2. Resource Recommendations' Examples

Based on waste needs and offers of our testing ecosystem, as described in Table 1, we describe two possible resolution scenarios, following the steps previous described.

4.2.1. "Wanting Wooden Packaging" Resolution Workflow

Supposing a query stated by the SweetHome administrative user, looking for wooden packaging, described as:

- *Keyword(s):* packaging
- *Direction:* Want
- *EWC:* 15 01 03

  ○ Search method: Filter

- Category: null
- *Site: SweetHome* geolocation

The algorithm first evaluates the direction of the query and the registered available resources in the system are filtered based on this attribute, returning only the resources of the opposite direction. In this case, it is a "want" resource. Therefore, the records of sites that "have" resources are retrieved and further evaluated.

As a second step, the similarity of the retrieved resources with the queried one is evaluated. In this example, the search method is set to "filter" therefore only resources with EWCs exactly matching the query EWC could be used. As here no resource category has been selected, this attribute is not being evaluated. From the existing resources, we see that the only "have" resources with the same EWC are

the wooden packaging offered by HotelNew and the wooden pallets that ChefRest has. For instance, the cardboard packaging offered by HotelNew has a EWC equal to 15 01 01 [48]. From those two candidates we now evaluate the linguistic descriptions written in the keywords text. As the query has been for "packaging" a 50% of the words in the resource's description of the HotelNew, i.e., wooden packaging, is matching the query, while the wooden pallets do not.

Therefore the recommended resource are the wooden packaging offered by HotelNew. If this recommendation is accepted and a synergy is established between the two parties, the conditions (like final quantity, price, date, etc.) can be negotiated between the two parties and the synergy included into the synergies case base for being considered in future. The added value of the recommender is to refine all possible packaging alternatives in the system that would be listed if the recommender would not be in place, to just suggesting the ones that maximize the resources that the user really is looking for.

4.2.2. "Wanting Plastic Packaging" Resolution Workflow

Supposing a query stated by the ChefRest administrative user, looking for plastic packaging, described as:

- *Keyword(s):* packaging
- *Direction:* Want
- *EWC:* 15 01 02

  ○ Search method: Rank

- Category: null
- *Site: ChefRest* geolocation

In this case, the algorithm again retrieves only the records of sites that "have" resources. As the search method is set to "rank" not only resources with EWCs exactly matching the query EWC are taken into account. From the existing resources, tagged as "have", there is none with exactly the same EWC. However there exist different types of packaging, like *the cardboard packaging* and *wooden packaging*, both offered by *HotelNew*. For instance, the *cardboard packaging* has EWC equal to 15 01 01 and the *wooden packaging* 15 01 03 [48]. As calculated using Equation (5), both of these EWCs have a similarity degree of approximately 0.45, while both linguistic descriptions match the query keyword at a 0.5 degree of similarity. Finally, the sites' distance between the query and the provider sites is calculated. Both candidates will be presented to the user, with the nearer one being presented first.

In this case, the final decision relies on the user, as no resource exactly matches the query or there is no resource that significantly outperforms the other that can be automatically identified.

5. Discussion

The two case studies presented illustrate major use of the recommender system and its advantages. The first case study shown the *recommender filtering mode*. In that mode, the resources recommended are filtered to have the same EWC than the one specified in the query. This behavior is the usual in Information Systems querying a database of resources. Main problem is that this querying systems are based on an *exact-matching method* to recover possible resources. The results on this kind of systems is that the user needs to know perfectly the resource is searching for, and the system only will provide with the answer of the resource searched with the same EWC (like *wooden packaging* and *wooden pallets* in the case study), if it is available according to the database of resources, or will provide an empty answer if there is no available with the exactly same EWC. This is a great drawback. There could be similar waste materials, which, even though they do not have the same queried EWC, are really similar to the resource searched in the query (like *cardboard packaging* in the case study). Thus, usual Information Systems will never get the other possible candidate resources. Furthermore, the *recommender system used in filter mode* has an advantage to classic database querying systems. In addition to filtering the

possible resources by the exact EWC, as it applies a similarity measure computation based on the linguistic description of the resources and the keywords enumerated in the query, the recommender can order and give a priority value to each possible candidate (with the same EWC), like in the case study, where the system gives more weight to the *wooden packaging* resource than the *wooden pallets*, because the *wooden packaging* is more similar to the searched resource because share the *packaging* word in the description of the resource. This means that the recommender, even filtering resources give a more useful list of suggestions than usual Information Systems offering a similar management of resources for setting industrial symbiosis synergies.

The recommender system shows its full potential and great advantage over traditional Information Systems, on the situations that are illustrated by the second case study. When the recommender is used in rank mode. In the *recommender ranking mode* the resources recommended are finally ranked, according to a similarity measure, which takes into account the degree of similarity between the EWCs, the similarity between the keywords in the query and the linguistic description of the resources, the distance between the location of the resource, and the site of the user, etc. Thus, it is using a *similarity-matching method*. This way, like in the second case study, all of the candidate resources are ordered according to the similarity measure (like *cardboard packaging* and *wooden packaging*), and the most suitable ones are on the first positions of the list. This way, the possibility to find the most useful resource is maximized. This is a great advantage against Information Systems based on the exact-matching approach offered by a database querying system. In addition, given a concrete query, the recommender is able to *make aware* to the user of other *unknown possible resources partially matching* the characteristics of the resource queried by the user, like the *cardboard packaging* resource in the first case study, if a ranking mode would have been used.

The great value that is added by this hybrid recommender system is that it provides the SHAREBOX platform with a very flexible recommendation system for suggesting the most suitable resources, given a query of a user, and on the other hand, provides it with a *resource discovering functionality*, with its hybrid recommendation strategy. As a consequence, the reusing of waste products, the reductions on new raw material for productive industrial processes, and at the same time, the energy cost savings that are derived on the established synergies are of great value.

The proposed synergies open the opportunity that the user introduces a by-product as an input to their processes, with the corresponding economic, energetic, and environmental savings, provided that the by-product will not need waste management treatments and the user will save all the costs of producing the corresponding primary resource. In the particular case that is presented in the paper, where all involved actors in the synergy are located in the same industrial park, there are, in addition, savings on transportation costs. In the midterm, it is expected that the feedback accumulated by the global synergy recommender will make the platform able to suggest unexpected synergies as well that the user a priori could not even be conceive and would never formalize in a query by himself or herself, thus contributing to the ecological innovation in industry.

6. Conclusions and Future Work

This paper describes an intelligent hybrid recommender system that is included in the SHAREBOX platform, which is able to identify relevant new by-products that might match the needs of a firm and can substitute primary resources. The main contribution of the paper is that the system is based on the integration of the Case-Based Reasoning approach and Ontological knowledge from the EWCs. The recommender system uses the EWC codes as a standard resources ontology that permits identifying semantic proximity between resources and improving the system performance in general. Specific similarity functions are defined that exploit the structure of the EWC ontology, thus including the semantics of the EWC codes themselves in the similarity computation. In addition, other features are taken into account to get the similarity between the textual descriptions of the resources/by-products and the keywords that were provided by the end user in the queries. As an additional innovation, the similarity between the user query and the cases in the Case Base are enriched with information

regarding a second classification of wastes, which provides an alternative ontology of waste categories and the geolocalization of the two involved firms to prioritize synergies between closer companies, thus increasing the feasibility of the synergy.

All these features provide the hybrid recommender system with a great flexibility, which enables and multiplies the probabilities to find the best resource/by-product options among the ones stored in the resource databases. This flexibility is a great advantage to conventional Information Systems, which offer the user exact-matching queries to the databases.

The assessment of the recommender system in the SHAREBOX platform has been satisfactory. Nonetheless, further evaluation and validation tests must be performed to confirm the very good preliminary evaluation of the system.

On the other side, a second level of recommendation can rely on the historical successful synergies that were performed in the SHAREBOX platform. This other synergy recommender system will increase performance as far as the platform accumulates new experience. The global synergy recommender is currently being integrated in the global platform and it can learn in the long term from the feedback regarding the success of the recommendations provided. This other component will make the platform more competent along time, because it will learn to recommend more synergies, and in a more accurate way.

References

1. Ricci, F.; Rokach, L.; Shapira, B. Introduction to recommender systems handbook. In *Recommender Systems Handbook*; Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Eds.; Springer: Boston, MA, USA, 2011; pp. 1–35.
2. Chertow, M.R. Uncovering Industrial Symbiosis. *J. Ind. Ecol.* **2007**, *11*, 11–30. [CrossRef]
3. Gatzioura, A.; Sànchez-Marrè, M.; Gibert, K. A Hybrid recommender system for industrial symbiotic networks. In Proceedings of the 9th International Congress on Environmental Modelling & Software (iEMSs 2018), Fort Collins, CO, USA, 24–28 June 2018.
4. Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Model. User Adapt. Interact.* **2002**, *12*, 331–370. [CrossRef]
5. Bistline, J.E.; Rai, V. The role of carbon capture technologies in greenhouse gas emissions-reduction models: A parametric study for the US power sector. *Energy Policy* **2010**, *38*, 1177–1191. [CrossRef]
6. European Communities. *Waste Management Options and Climate Change*; European Communities: Brussels, Belgium, 2001.
7. Climate Change. *United Nations Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2001; 3 Volumes. Available online: www.ipcc.ch (accessed on 13 September 2019).
8. European Environmental Agency. *More from Less Material Resource Efficiency in Europe*; Technical Report, EEA Report No. 10/2016; European Environmental Agency: Copenhagen, Denmark, 2016; Volume 10.
9. Walmsley, T.G.; Ong, B.H.; Klemeš, J.J.; Tan, R.R.; Varbanov, P.S. Circular Integration of processes, industries, and economies. *Renew. Sustain. Energy Rev.* **2019**, *107*, 507–515. [CrossRef]
10. Klemeš, J.J.; Kravanja, Z. Forty years of Heat Integration: Pinch Analysis (PA) and mathematical programming (MP). *Curr. Opin. Chem. Eng.* **2013**, *2*, 461–474. [CrossRef]
11. El-Halwagi, M.M. *Sustainable Design through Process Integration*; Butterworth-Heinemann/Elsevier: London, UK, 2012.
12. Porzio, G.F.; Colla, V.; Matarese, N.; Nastasi, G.; Branca, T.A.; Amato, A.; Bergamasco, M. Process integration in energy and carbon intensive industries: an example of exploitation of optimization techniques and decision support. *Appl. Therm. Eng.* **2014**, *70*, 1148–1155. [CrossRef]
13. Kermani, M.; Kantor, I.D.; Wallerand, A.S.; Granacher, J.; Ensinas, A.V.; Maréchal, F. A Holistic Methodology for Optimizing Industrial Resource Efficiency. *Energies* **2019**, *12*, 1315. [CrossRef]

14. The Biofuels Directive (2015/1513). European Parliament, 2015. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015L1513&from=EN (accessed on 13 September 2019).

15. European Commission. *Renewable Energy Directive 2009/28/EC*; European Commission: Brussels, Belgium, 2009.

16. Eurostat. *Energy Balance Sheets 2011–2012*; Technical Report 9; Eurostats (European Union): Brussels, Belgium, 2014. [CrossRef]

17. EC Directive 1999/31/EC. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A31999L0031&from=EN (accessed on 13 September 2019).

18. EC Directive 2000/53/EC on End-of-Life Vehicles. Available online: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2000L0053:20050701:EN:PDF (accessed on 13 September 2019).

19. EC Directive 2002/96/EC on Waste Electrical Equipment. Available online: https://ec.europa.eu/environment/waste/weee/index_en.htm (accessed on 13 September 2019).

20. EC Directive 2006/66/EC on Batteries. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0066&from=EN (accessed on 13 September 2019).

21. Pearce, D.W.; Turner, R.K. *Economics of Natural Resources and the Environment*; JHU Press: Baltimore, MD, USA, 1990.

22. Frosch, R.A.; Gallopoulos, N.E. Strategies for Manufacturing. *Sci. Am.* **1989**, *261*, 144–152. [CrossRef]

23. McKinsey & Company. *Towards the Circular Economy: Economic and Business Rationality for an Accelerated Transition*; Ellen McArthur Foundation: Cowes, UK, 2012.

24. EUR-Lex. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0190&from=EN (accessed on 13 September 2019).

25. European Commission. *EC Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Implementation of the Circular Economy Action Plan*; European Commission: Brussels, Belgium, 2019.

26. Lombardi, D.R.; Laybourn, P. Redefining Industrial Symbiosis. *J. Ind. Ecol.* **2012**, *16*, 28–37. [CrossRef]

27. Chertow, M.R. Industrial Symbiosis: Literature and Taxonomy. *Ann. Rev. Energy Environ.* **2000**, *25*, 313–337. [CrossRef]

28. Costa, I.; Massard, G.; Agarwal, A. Waste management policies for industrial symbiosis development: case studies in European countries. *J. Clean. Prod.* **2010**, *18*, 815–822. [CrossRef]

29. Van Capelleveen, G.; Amrit, C.; Yazan, D.M. A literature survey of information systems facilitating the identification of industrial symbiosis. In *From Science to Society*; Springer: Cham, Switzerland, 2018; pp. 155–169.

30. Van Capelleveen, G.; Amrit, C.; Yazan, D.M.; Zijm, H. The influence of knowledge in the design of a recommender system to facilitate industrial symbiosis markets. *Environ. Model. Softw.* **2018**, *110*, 139–152. [CrossRef]

31. Yazdanpanah, V.; Yazan, D.M.; Zijm, W.H.M. FISOF: A formal industrial symbiosis opportunity filtering method. *Eng. Appl. Artif. Intell.* **2019**, *81*, 247–259. [CrossRef]

32. Deshpande, M.; Karypis, G. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.* **2004**, *22*, 143–177. [CrossRef]

33. Melville, P.; Sindhwani, V. Recommender Systems. In *Encyclopedia of Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2010; chapter: 00338; pp. 829–838.

34. Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 734–749. [CrossRef]

35. Su, X.; Khoshgoftaar, T.M. A Survey of Collaborative Filtering Techniques. *Adv. Artif. Intell.* **2009**, *2009*, 1–19. [CrossRef]

36. Lopez de Mantaras, R. Case-Based Reasoning. In *Machine Learning and Its Application*; LNAI; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2049, pp. 127–145.

37. Kolodner, J.L. An introduction to case-based reasoning. *Artif. Intell. Rev.* **1992**, *6*, 3–34. [CrossRef]

38. Bridge, D.; Göker, M.H.; McGinty, L.; Smyth, B. Case-based recommender systems. *Knowl. Eng. Rev.* **2006**, *20*, 315. [CrossRef]

39. Leake, D.B. *CBR in Context: The Present and Future. Case-Based Reasoning: Experiences, Lessons, and Future Directions*; MIT Press: Cambridge, MA, USA, 1996; pp. 3–30.

40. Núñez, H.; Sànchez-Marrè, M.; Cortés, U.; Comas, J.; Martínez, M.; Rodríguez-Roda, I.; Poch, M. A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations. *Environ. Model. Softw.* **2004**, *19*, 809–819. [CrossRef]

41. Finnie, G.; Sun, Z. Similarity and metrics in case-based reasoning. *Int. J. Intell. Syst.* **2002**, *17*, 273–287. [CrossRef]

42. Gibert, K.; Nonell, R.; Velarde, J.M.; Colillas, M.M. Knowledge Discovery with clustering: impact of metrics and reporting phase by using KLASS. *Neural Netw. World* **2005**, *4*, 319–326.

43. Gibert, K.; Valls, A.; Batet, M. Introducing semantic variables in mixed distance measures. Impact on hierarchical clustering. *Knowl. Inf. Syst.* **2014**, *40*, 559–593. [CrossRef]

44. Lorenzi, F.; Ricci, F. Case-based recommender systems: A unifying view. In *Intelligent Techniques for Web Personalization*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 89–113.

45. Gatzioura, A. A Hybrid Approach for Item collection Recommendations: An Application to Automatic Playlist Continuation. Ph.D. Dissertation, Universitat Politècnica de Catalunya, Barcelona, Spain, 2018.

46. Gatzioura, A.; Sànchez-Marrè, M. A case-based reasoning framework for music playlist recommendations. In Proceedings of the 4th IEEE International Conference on Control, Decision and Information Technologies (CoDIT'17), Barcelona, Spain, 5–7 April 2017.

47. Scottish Environment Protection Agency. Available online: https://www.sepa.org.uk/media/139146/sepa-waste-thesaurus-2014.pdf (accessed on 1 August 2019).

48. National Waste Collection Permit Office. Available online: http://www.nwcpo.ie/forms/EWC_code_book.pdf (accessed on 1 August 2019).